Discussion Questions
FFDM Radiological Devices Panel Meeting
November 17, 2009

1. The FFDMs that have been approved through PMA applications have indications for both screening and diagnostic mammography, based on clinical studies that were designed to assess screening performance of FFDM compared to film screen mammography. In your opinion, is screening performance indicative of diagnostic performance, or should new FFDMs have separate testing supporting screening and diagnostic indications, if they are indicated for both?

2. The draft guidance describes numerous physical laboratory tests to be conducted on full field digital mammography systems.

   Can physical laboratory testing alone provide sufficient information about safety and effectiveness of FFDM performance to be the only testing required to demonstrate substantial equivalence of a new FFDM? If not, what are the limitations of this approach?

3. If clinical testing is needed in addition to laboratory testing to demonstrate substantial equivalence, what should be the purpose of the clinical study? Please consider the following possibilities.

   i. To ensure that there are no unanticipated problems in the imaging system.

   ii. To provide user preference information that may be conveyed in labeling.

   iii. To provide a comparison of the screening and diagnostic performance of the device including the following:
      - detect microcalcifications
      - discriminate benign from malignant microcalcifications
      - detect regions of architectural distortion
      - discern subtle irregularities in otherwise smooth mass margins and thereby discriminate between benign and malignant masses.

   iv. Other

4. An assumption in the guidance document is that a device that performs well with easy or normal cases might not be adequate for difficult cases; hence difficult cases must be assessed to evaluate differences in performance (stress testing).

   a. Do you agree that a device that performs well with easy or normal cases might not perform well with difficult cases?

   b. Do you believe that stress testing is necessary for an adequate clinical testing?

c. Do you agree with the recommendation in the draft guidance that mammograms collected for a stress study have the following characteristics? If not, what do you recommend?
   - all patient lesions less than 1.0 cm in size and non-palpable
   - all breast compositions but predominantly dense (i.e., at least 75%), with an equal number of heterogeneously dense and homogeneously dense
   - even distribution of masses, clusters of microcalcifications, and architectural distortions (majority malignant but a sufficient number benign)
   - a distribution of clusters of microcalcifications that includes different types of benign and malignant microcalcifications
   - at least one retroareolar mass, one retroareolar cluster of microcalcifications, and one retroareolar architectural distortion
   - a small number of normal mammograms.

5. If clinical data are needed to demonstrate substantial equivalence, there are different approaches that may be followed to acquire the necessary information. Please discuss the benefits and drawbacks of the following approaches, and whether there are different conditions when different approaches would be preferred.

   a. <u>ACR-like model:</u> A free standing analysis of about 30-cases from the new FFDM only to assess image quality.

   b. <u>Mammographic Feature Analysis:</u>

      i. Should the cases be selected to stress the system?

      ii. The draft guidance document lists 14 features that could be compared. Are all of these equally important, or is there a smaller number of features that are most important (e.g. lesion conspicuity and positioning)? What features do you recommend be compared in order to demonstrate substantial equivalence?

      iii. What should the endpoints of a features analysis be? Is it acceptable to use a "small" set of images acceptable (e.g. 30) and look at trends, or should the study be sized for statistical significance of one or more endpoints?

      iv. Should the cases be paired (double exposure of patients), or should there be different datasets for the new FFDM and the predicate? If there are separate datasets, should equivalence of the datasets be demonstrated, and, if so, how?

   c. <u>MRMC study:</u> A multiple reader multiple case (MRMC) study designed to show non-inferiority of the new device with the predicate in one or more performance metrics (e.g., sensitivity, specificity, area under the ROC curve (AUC), or partial AUC).

6. Given the above discussions, what is your recommendation for the minimum data needed to establish the substantial equivalence of a new FFDM device to a predicate?